图书情报工作
第63卷 第9期 2019年5月

社交媒体用户画像的构建及资源聚合模型研究

■ 徐海玲[1] 张海涛[1,2] 魏明珠[1] 尹慧子[1]

[1] 吉林大学管理学院 长春 130022 [2] 吉林大学信息资源研究中心 长春 130022

摘要：[目的/意义]基于用户画像创建社交媒体资源聚合模型,为资源的聚合提供参考,丰富和拓展其理论研究体系。[方法/过程]在深度剖析用户画像的内涵、算法的基础上,基于社交媒体分别构建用户画像和资源画像的模型,运用社会化标签系统方法,研究基于社交媒体的用户画像和资源画像之间的映射关系。借鉴领域本体的方法,通过对数据的深层次挖掘,利用资源聚合的原理,构建基于资源画像的社交媒体内部资源聚合模型和外部资源聚合模型。[结果/结论] 在大数据时代,基于用户画像和资源画像的相关理论与方法,能够为社交媒体的资源聚合提供新的思路。

关键词：用户画像 社交媒体 资源画像 资源聚合

分类号：G252

DOI：10.13266/j.issn.0252-3116.2019.09.012

1 引言

随着物联网、云计算、大数据、人工智能领域时代的到来,社交平台已经得到了人们的高度依赖。在开放的环境下,用户不仅是资源的使用者,同时也是资源的创造者和资源的共享者。随着用户信息行为轨迹更易捕捉,用户的数据更易获取,根据用户的需求,为用户提供精准、客观、动态的服务变得越发重要,已经成为学者们关注的前沿和热点。随着社交网络的兴起,微博、微信、豆瓣等社区得到了快速的发展,通过社交网站,用户不仅可以感知天下大事,与此同时还可以足不出户查找感兴趣的影音讯息。由于社交媒体用户人数逐年增多,满足用户多样化、个性化的服务将成为人们追求的目标。鉴于此,笔者尝试在吸收和借鉴前人经验和结果的基础上,对社交媒体的资源特征进行分析。通过深入挖掘用户的行为数据,从而进行社交媒体的用户画像与资源画像的构建。在整合用户行为构成要素的基础上,搭建社交媒体的资源聚合模型,主要分为内部资源聚合模型和外部资源聚合模型,为用户信息资源聚合和推送的研究与应用提供一种新的思路和尝试。本文的逻辑思路见图1。

2 文献回顾

随着人工智能、大数据领域的悄然兴起,用户画像领域的研究才逐步成为众多学者关注的热点,并取得了一定的研究成果。杨帆[1]从读者画像和资源画像着手,构建了图书馆的大数据平台,为图书馆用户提供精准的服务;王顺箐[2]通过对图书馆用户进行需求分析,基于用户的兴趣爱好为他们提供差异化的推荐服务;尹相权等[3]通过对图书馆用户的数据进行建模分析,得出了影响用户行为的主要因素;K. Petric 等[4]基于自适应的知识管理方法,开发了一个网络用户的心智模型,实现了对不同的用户群体进行分类;程全[5]探讨了基于用户画像的数字图书馆信息服务的基本模式,为数字图书馆信息的精准服务提供科学决策;王凌霄等[6]从4个方面(用户资历、用户参与度、用户回答质量和用户发展趋势)构建了社会化问答社区用户画像;单晓红等[7]从用户信任属性、酒店信息属性和用户评价信息属性3个方面构建了携程酒店用户画像的概念模型;黄文斌等[8]从移动数据的类型、移动用户行为模式、移动用户画像的建构和移动用户画像的深度应用等几个方面构建了数据驱动的移动用户行为的研究框架;郝胜宇等[9]从目标顾客识别、目标顾客扫描、消费异动和精准推送等几方面详细论述了用户画像技术,以及用户画像对企业精准营销的重要作用;裴国才[10]通过用户画像的方法,设计并实现了精准营销的模型;王晓霞等[11]利用大数据技术对用户进行精准聚类从

作者简介：徐海玲(ORCID:0000-0001-6887-778X),博士研究生,E-mail:269070496@qq.com;张海涛(ORCID:0000-0002-9421-8187),教授,博士生导师;魏明珠(ORCID:0000-0001-8430-7461),硕士研究生;尹慧子(ORCID:0000-0002-4281-1333),博士研究生。

收稿日期:2018-07-06 修回日期:2018-10-19 本文起止页码:109-115 本文责任编辑:徐健

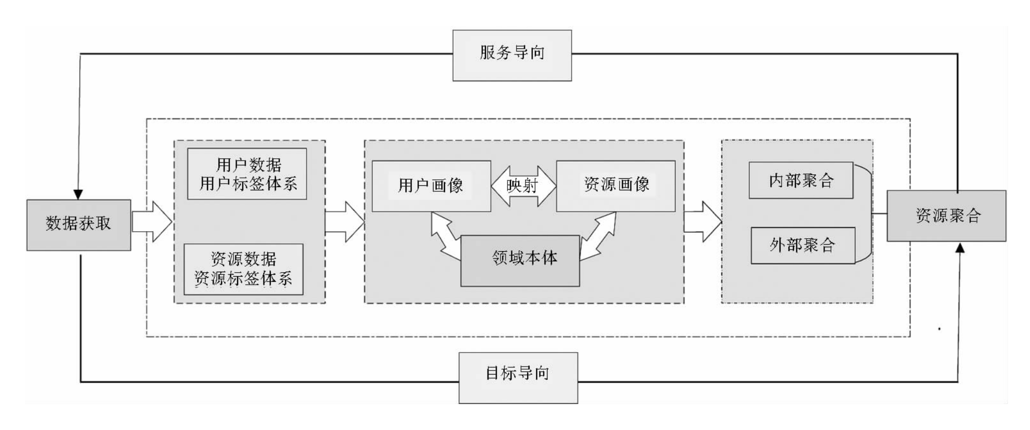


图 1　本文逻辑思路

而刻画用户画像;盛怡瑾[12]将用户画像技术应用到学术期刊审稿人遴选中,并构建了审稿人画像模型;熊伟等[13]通过采集用户信息并对用户群体进行分类,用 LDA 分析网页的具体内容建立主题模型,提出了一种基于用户画像和内容的服务重定向方法;张诗军等[14]基于用户画像的理论与技术提出了电力大数据的客户画像构建方法;高扬等[15]对智能制造领域的人才画像进行了构建,揭示了该领域杰出人才的显著特征。

经过对相关文献的梳理可以看出,目前用户画像的研究领域较为宽泛,多集中于图书馆学、社交媒体、计算机以及市场营销领域,研究方法较为单一,多以实证研究为主。因此,笔者尝试在吸收和借鉴前人研究成果的基础上,基于社交媒体这一平台,实现用户画像和资源画像的刻画,并建立二者之间的映射关系,着力构建平台内部的资源聚合模型,同时打造跨平台的不同社交媒体之间的资源聚合模型,为用户智能终端的信息资源推送提供参考,为该领域的相关研究提供理论支撑。

3　用户画像的研究概述

3.1　用户画像的内涵

最早提出用户画像(persona)的学者是被命名为交互设计之父的阿兰·库珀,该学者认为,用户画像是真实反映用户数据特征的虚拟代表,通过对用户数据的挖掘,用户的目标、行为和观点进行抽取,分析出用户的典型特征,把用户的静态数据和动态数据标签化,从而形成的一个目标用户的模型。用户画像的核心工作就是通过人为规定高度精确的特征标识来为用户打标签,主要目的就是通过标签,使计算机能够程序化地处理一些与人相关的数据信息,提高信息获取的速率,在还原用户信息的基础上,构建用户画像,从而为广告投放、市场营销等领域的信息推广提供服务。随着用户行为轨迹的更易捕捉,为用户提供精确、动态的信息变得越来越重要,用户画像的应用范围也在不断拓展,从最传统的营销领域的用户画像(persona)开始向社交媒体等领域用户画像(user profiling)逐步延伸。

3.2　用户画像研究主题、理论模型

通过对用户画像的相关论文进行脉络梳理可以发现,目前用户画像的研究主题大致分为以下几种:微博[16]、移动图书馆[17]、企业营销、社会化问答社区[6]、电子商务[18]、医疗领域和金融行业等;用户画像的模型主要包括:马尔可夫随机场模型、逻辑回归模型;用户画像的理论主要包括:社会认同理论[16]、复杂适应系统理论和博弈论等;用户画像的相关算法主要包括:层次分析法[19]、聚类算法[20]、支持向量机、朴素贝叶斯分类[21]、相关分析法、决策树分析法、神经网络分析法[22]等,如表 1 所示:

表 1　用户画像相关研究的主题模型

研究主题	理论模型	机器学习算法
微博、数字图书馆、社交媒体、社会化问答社区、网络安全、医疗领域、金融行业、电子商务	模型:马尔可夫随机场模型、逻辑回归模型 理论:社会认同理论、复杂适应系统理论、博弈论	层次分析法、聚类算法、支持向量机、朴素贝叶斯分类法、k-means 算法、神经网络算法、相关分析法、决策树分析法、BP 神经网络分析法

3.3　用户标签体系

用户画像的标签体系,就是用户信息的标签化,在构建用户画像之前,需要建立标准的标签体系,来全方位、多层次地反映出用户画像的基本内容。用户画像标签的建立,需要对数据进行采集和处理,基于数据的不同,以及需求的不一致,在采集标签时,大致采用两种方法:①基于人工手动归纳并标注用户的标签;②运用机器算法进行半人工的方式提取。标签指标体系的完善程度,对于用户画像的精准构建将起到至关重要的作用。在标签体系中,每一个标签都是某一个用户

特征的具体体现，从某种程度上来说，该标签要具有一定的群体性，能够归纳出事物的某种属性和基本的特征，标签在表现形式上多种多样，可以是中文的符号也可以是数字。从原始数据的收集和处理，到业务标签的生成，从数据的清洗和整理，到最终的机器学习，通过对用户的属性特征进行分类，都能够运用计算机进行处理[17]，如图2所示：

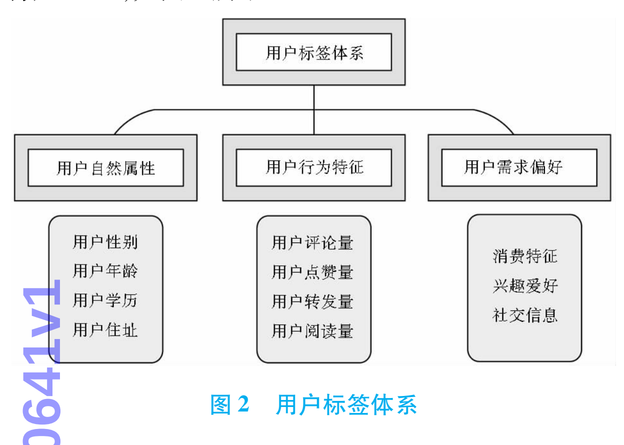


图2　用户标签体系

4　社交媒体的用户画像与资源画像构建

随着近年来，互联网产业的异军突起，各类型的社交媒体也如雨后春笋般蓬勃发展，对于社交媒体而言，为用户提供精准、个性化的推送服务攸关各大社交媒体的长远发展。豆瓣网是Web2.0时代最具代表性的网站之一，其在线功能较为丰富，豆瓣网已经由最初的提供读书、电影和音乐信息的网站，发展成一个集博客、商业为一体的综合性的社交网络平台，时至今日，豆瓣电影已经成为中国最大最权威的电影分享和在线评论的电影社区，甚至被称为华语圈的“IMDB”，目前有将近3 000多家电影院加盟其中，豆瓣电影已经成为豆瓣网络主网站点击流量的第一子模块。因此，笔者以豆瓣电影网为例，对豆瓣网的用户画像和资源画像进行构建，以期通过画像之间的映射关系，来实现个性化的推送服务[23]。

4.1　社交媒体用户画像构建

4.1.1　数据的获取　以“豆瓣”电影为例，运用八爪鱼爬虫软件，对豆瓣电影网2017－2018年热门电影进行了数据的爬取工作，爬取时间为2018年5月7日，经过对数据的筛选、清洗和加工，导出数据文件到Excel中，共爬取出电影数据221条，用户的评论数据为7 000条。将这些数据进行归类、汇总，对数据进行统计分析，爬取的内容主要归类为两大部分：近两年热门电影的相关数据和用户的相关数据。用户数据主要包括作品名称、作品ID、评论标题、评论者的ID、星级评分、推送程度、评论时间、评论摘要、评论内容、评论有用数、评论无用数、追评时间、追评内容等。资源数据主要包括电影片名、电影导演、电影编剧、电影主演、电影类型、电影制片国家、上映日期、电影评分、电影评论人数。

4.1.2　用户画像标签体系的建立　笔者基于用户数据，对用户的标签体系进行了划分，主要分为：用户的自然属性、用户的行为特征属性和用户的需求属性3个方面。用户的自然属性主要包括用户的ID、用户所在城市；用户的行为属性主要包括用户评分、用户推荐、用户评论和用户分享；用户的需求属性主要包括影片演员、影片导演、影片类型、影片上映档期、影片语种、评论得分和影片排行榜等，如图3所示：

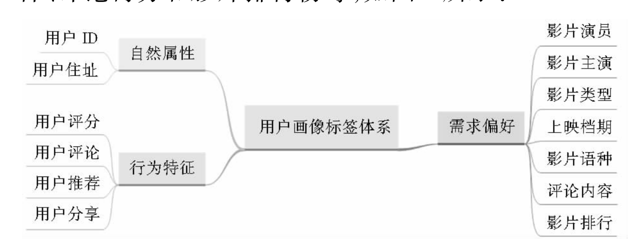


图3　豆瓣电影用户画像的标签体系

4.1.3　用户画像的构建　为了能够更直观地显示用户的标签体系，笔者将易词云软件应用于用户画像的构建中，用户的标签主要以用户的行为特征和需求特征为研究的核心，在词云中，字体越大，代表该类用户的标签处于核心地位，并且在用户的标签体系中占有着决定性的地位，相反，字体越小，代表该类用户的标签处于非核心地位，作用相对较弱。笔者以豆瓣网的用户群体为例，构建的用户画像如图4所示：



图4　豆瓣电影用户画像词云图

4.2　社交媒体资源画像构建

4.2.1　资源画像标签体系的建立　资源画像的标签体系大致分为3个层级，分别是影片的自然属性、影片的特征属性和影片的内容属性，影片的自然属性主要包括影片的片名、影片的上映年份、影片的导演；影片

的特征属性主要包括影片的评价得分、影片的评价人数、影片的影评、影片的媒体宣传、影片的话题宣传和影片的路演宣传；在影片的内容分类里，大致分为两大部分，分别为影片对人的需求标签体系，以及影片本身的内容标签体系。影片对人的需求标签，主要包括影片的价格、影院的环境、影片的类型、影片的语种；影片内容标签主要包括影讯、影片购票、影片排行榜、影片分类、影评、影片年度榜单和观影报告等[24]。如表 2 所示：

表 2 豆瓣电影资源画像标签体系

项目	内容	
影片的自然属性	影片的片名	
	影片的上映年份	
	影片的导员、演员和编剧	
影片的基本特征	影片的评价得分	
	影片的评价人数	
	影片的影评	
	影片的媒体宣传、影片的话题宣传和影片的路演宣传	
影片的服务内容	人的需求标签	影片的价格
		影院的环境
		影片的类型
		影片的语种
	影片的内容标签	影讯
		影片购票
		影片排行榜
		影片分类
		影评
		影片年度榜单
		观影报告

4.2.2 资源画像的构建 运用易词云软件，基于资源画像的标签体系，对资源画像进行了构建，由于资源画像的标签体系较为复杂，所以剔除了资源标签中重复的数值，并将资源的标签单词数值设置为 1，生成数据格式，并将单词的间隔设置为 0，使生成数据的单词间隔间隙较适中，构建的资源画像如图 5 所示：



图 5 豆瓣电影资源画像词云图

4.3 用户画像与资源画像之间的映射关系

社会化标签系统是由 3 种不同类型的集合构成的，分别为用户、资源和标签，它们三个之间形成了一个完整的大众分类法（folksonomy）。在该系统中，每一组数据之间都遵循着 F =｛用户、资源，标签 1，标签 2，……标签 t｝，F =｛user，item，tag1，tag2……tagt｝，其中，t 是由用户分配给不同资源的标签的数量。

定义 $M = \{M_1, M_2, M_3, M_4, \cdots\cdots M_i\}$、$V = \{V_1, V_2, V_3, V_4, \cdots\cdots V_j\}$、$T = \{T_1, T_2, T_3, T_4, \cdots\cdots T_k\}$，其中，M、V 和 T 分别代表三个不同的数据集，M 代表用户集、V 代表资源集、T 代表标签集。i、j、k 分别代表各自对应数据集的数量[25]。

同时，每一组的关系有相应的矩阵构成，具体有 B，B′，B″，分别代表着用户与资源、资源与标签以及用户与标签。在矩阵 B 中，如果 M_i 选择了资源 T_k，则 $b_{ik} = 1$，否则 $b_{ik} = 0$。同理可得，在 B′矩阵中，如果 Vj 被标记了 T_k，那么 $b_{jk} = 1$，否则 $b_{jk} = 0$。在 B″矩阵中，如果 M_i 选择了 V_j，则 $b_{ij} = 1$，否则 $b_{ij} = 0$。

笔者借鉴社会化标签系统的模型，将该模型中 M 代表用户画像，V 表示资源画像，对用户画像和资源画像二者之间建立 T 的标签集合，从而建立了用户画像与资源画像之间的映射关系，如图 6 所示：

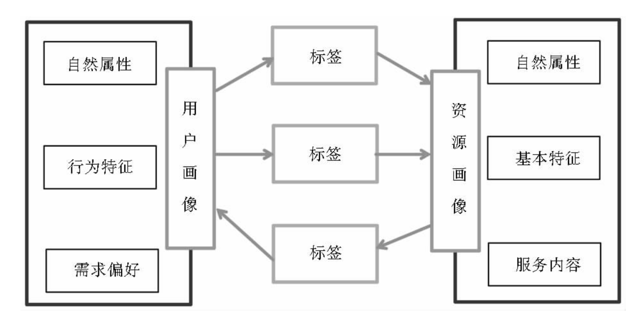


图 6 用户画像与资源画像之间的映射关系

在该映射关系中，通过在用户画像与资源画像之间建立标签，使它们之间的映射关系得以成立。其中，M =｛用户的自然属性，用户的行为属性，用户的需求属性｝、V =｛影片的自然属性，影片的特征属性，影片的内容属性｝、T =｛行为、需求、特征、内容｝。利用聚类算法将标准标签作为聚类的中心，计算用户画像和资源画像的自定义标签与聚类中心标签的相关度，将自定义标签聚集到相关度最大的聚类中心的所属类中。直到剩余的自定义标签与聚类中心的相关度均小于阀值，停止聚类，将剩下的自定义标签互相聚类，将新的聚类中心存入到标准的标签库中，重新计算自定义标签与标准标签的相似度，并调整自定义标签的聚类，得到新的标签的标准聚类。通过该种方法实现用户画像与资源画像之间的映射[26]。笔者以豆瓣电影

网站为例，将豆瓣电影网站中的用户画像和资源画像的属性进行具体细分，并建立标签，形成了具体的用户画像映射模型，如图 7 所示：

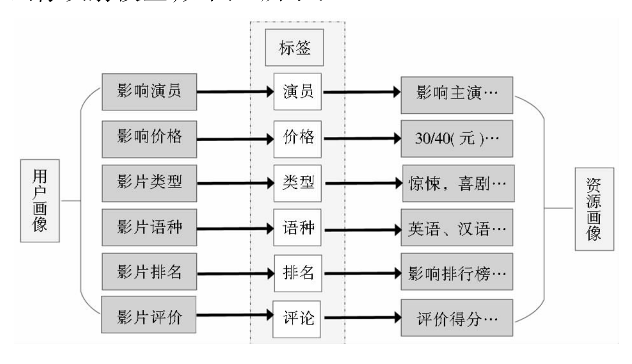


图 7　用户画像与资源画像的映射模型

在该模型中，依据用户评论、点赞等信息，帮助用户查找感兴趣的电影资源。通过将电影的关键词汇和关系存储在标签体系中，依据用户的检索信息，在标签栏中进行词汇的高级匹配，使资源画像的检索内容与检索关键词相对应，为用户实现推送服务，这种方式在某种程度上，不仅满足了用户个性化的需要，而且提高了推送的服务质量。例如：用户在标签中输入排名，这时检索系统会根据用户的检索需求，在资源画像中会呈现影片的年度榜单和影片的排行榜等信息，从而能够为用户提供精准的推送。

5　社交媒体的资源聚合模型构建

5.1　模型的构成要素

资源聚合是指对互联网上多种资源的特征信息或者不同类型的数据信息进行搜集和筛选、分析和整理，通过对资源的数据挖掘来分析资源之间的潜在关系，根据资源之间的属性关系实现资源的聚集，从而为用户提供知识资源与信息的过程。

5.1.1　用户画像与资源画像　用户是社交媒体资源聚合和服务推送的主体，是社交媒体资源的利用者，主要指参与者在社交媒体中进行知识的交流、分享、传递形成的各种类型的资源聚合形式。用户画像实现了对用户群体的划分，将相似的特征通过整合的形式划分为一类，并将其聚集起来，实现用户的群体推送和个性化推送服务。社交媒体中的资源主要包括社交媒体中以数字化形式呈现的各种信息化的资源聚合，这里不仅包括我们所熟悉的文字、图片、音频、视频和广播，还包括社交媒体中无形资源，如媒体的价值观和情感资源等。社交媒体中的资源以多样化的形式存在，通过资源画像将这些资源整合，达到全面记录所需的知识以及为用户表达所需知识资源的目的。

5.1.2　领域本体　领域本体是指概念与概念之间的映射关系，由本体构建者的需求决定，可以是一个学科领域或者是某几个领域的知识的结合，也可以是领域中的一个范围。随着社交媒体领域资源的不断聚合，组织结构也发生了显著的变化，由原来的单一的线性的组织结构向多维度、多空间的网状结构发展，社交媒体的资源聚合形式也客观地体现了媒体之间资源的关联关系，从现有的研究领域可以看出，当前领域本体的研究已经不仅是原有知识的传统的属性结构，更多的是概念和属性关联关系的多重集成，从而使领域本体关注的对象也发生了显著的变化，从单一的主题向相似主题、多维度主题转化，语义的信息已经向概念之间的映射与关联关系方面不断拓展，多元化信息资源之间的聚合关系逐渐显现[27]。

5.2　资源聚合模型

资源聚合是为了解决资源孤岛、信息超载和冗余等问题以实现资源的快速获取。笔者基于用户画像的研究概述、用户画像的模型构建，以及用户画像和资源画像之间的映射关系，在对资源聚合的概念进行分析的基础上，构建了资源画像的资源聚合模型，由于资源分布的广泛性，以及用户群体需求的多样性，因此在对资源进行聚合时，笔者着力打造两种资源聚合模型，分别为基于社交媒体资源画像的内部资源聚合模型以及实现跨平台的基于不同社交网络媒体资源画像的外部资源聚合模型，从而为用户提供精准的服务。

5.2.1　基于社交媒体资源画像的内部资源聚合模型　内部聚合主要是针对社交媒体内部的用户画像、资源画像和本体领域的聚合。通过对繁杂、无序资源之间的关系进行深度挖掘，建立特征之间的联系，并以用户画像为基础，构建画像之间的映射关系从而达到资源聚合的目的。社交媒体资源画像的内部资源聚合模型，见图 8。

资源画像内部资源聚合的模型主要分为三个部分，分别为资源层、算法层和画像层。资源层是资源聚合和服务的推送基础，在数据库中，主要包括豆瓣电影用户的数据和豆瓣电影网站资源的数据，其中电影的资源数据主要包括电影、图片、音频、视频和广播等数字资源，用户的数据主要包括用户的基本信息、行为信息和需求信息等，通过对数据信息的整理和加工，获取所需要的资源；算法层是通过对豆瓣电影网站用户的数据进行清洗和处理，由于其信息的概念粒度较细，应用机器学习的方法对用户和资源的数据进行深层次的

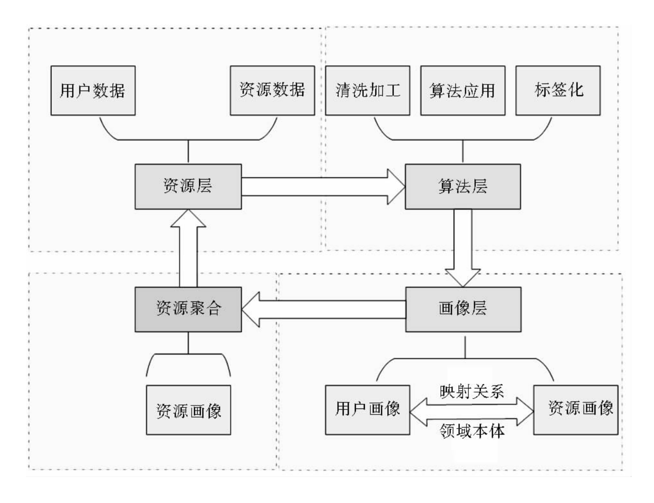


图 8 基于社交媒体资源画像的内部资源聚合模型

加工，挖掘其本质的内涵和信息，对豆瓣电影中的资源进行再组织，从诸多的数据中抽取出资源的标题，特征和关键词，对电影的资源进行特点的分类，从而建立标签体系；画像层是根据用户画像和资源画像之间的映射关系，基于标签形成三者之间的相似度矩阵，寻求本体之间的关联关系，根据用户的需求，为用户提供精确的匹配结果，通过资源画像实现资源聚合，为用户提供有价值的推送信息，从而实现良好的推送服务。

5.2.2 基于社交媒体资源画像的外部资源聚合模型

在电影网站的外部，分布着不同的社交电影媒体的平台，它们之间既存在竞争，同时也存在着合作，因此，在资源的聚合方面，要实现跨平台信息资源的聚合，从而达到为用户提供更精准服务的目的。基于社交媒体资源画像的外部资源聚合模型，如图 9 所示：

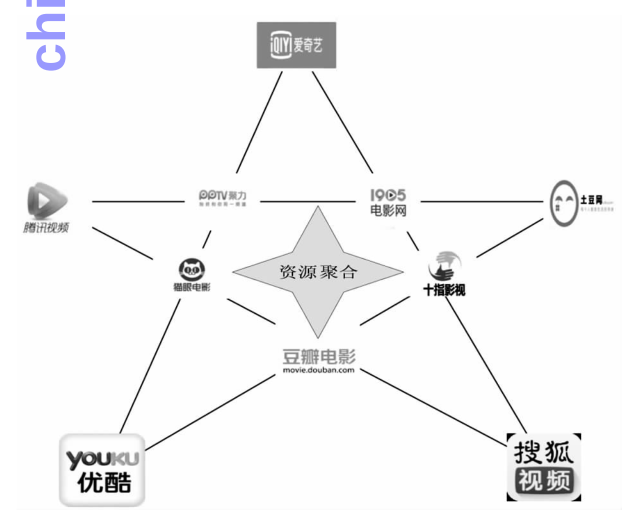


图 9 基于社交媒体资源画像的外部资源聚合模型

通过图 9 可以看出，在豆瓣电影网站的外部存在着诸多电影类网站，如优酷、土豆、乐视、爱奇艺、搜狐视频和腾讯视频等。这些网站与豆瓣网站一样，存在着大量的用户群体和海量的电影资源，因此，通过分析豆瓣电影的资源信息以及资源画像，可以实现资源之间的匹配和聚集，达到不同资源网站之间的交流和合作，从而为用户提供跨平台的资源推送服务。

6 结论

笔者基于用户画像的内涵、理论和算法的相关研究，对社交媒体的用户画像进行了深层次的剖析，并以豆瓣网为例，提炼汲取了基于社交媒体用户画像和资源画像的基本模型，并基于社会化标签系统的基本方法，厘清了用户画像、标签和资源画像之间的映射关系，构建了映射关系模型。结合资源聚合的内涵、领域本体的相关知识，深入解析了基于社交媒体的资源画像资源聚合的影响因素，构建了基于资源画像的社交媒体内部资源聚合模型和外部资源聚合模型，为具有针对性的信息推送与服务奠定基础。希望笔者的研究能够为本领域相关研究提供一定的理论基础，进而丰富和完善用户画像的相关研究。本文的研究结论如下：用户画像是真实反映用户数据特征的虚拟代表，通过对用户数据的挖掘，分析出用户的典型特征，从而形成一个目标用户的模型。基于社交媒体的资源画像资源聚合模型主要分为内部资源聚合模型和外部资源聚合模型，通过两种模型的运用，来为用户提供个性化的服务。

参考文献：

[1] 杨帆. 以画像分析为基础的图书馆大数据——以国家图书馆大数据项目为例[J]. 图书馆论坛, 2018(4): 1－9.
[2] 王顺箐. 以用户画像构建智慧阅读推荐系统图书馆学研究[J]. 图书馆学研究, 2018(4): 92－96.
[3] 尹相权, 李书宁, 弓建华. 基于系统日志的高校图书馆研究间用户利用行为分析[J]. 现代情报, 2018(1): 115－120.
[4] PETRIC K, PETRIC T, KRISPER M. User Profiling on a Pilot Digital Library with the Final Result of a New Adaptive Knowledge Management Solution[J]. Knowledge organization, 2011, 38(2): 96－113.
[5] 程全. 基于用户画像的数字图书馆信息服务模式研究[J]. 图书馆学刊, 2018(4): 68－71.
[6] 王凌霄, 沈卓, 李艳. 社会化问答社区用户画像构建[J]. 情报理论与实践, 2018(1): 129－134.
[7] 单晓红, 张晓月, 刘晓燕. 基于在线评论的用户画像研究——以携程酒店为例[J]. 情报理论与实践, 2018(4): 99－104, 149.
[8] 黄文斌, 吴家辉, 徐山川, 等. 数据驱动的移动用户行为研究框架与方法分析[J]. 情报科学, 2016(7): 14－20, 40.
[9] 郝靖宇, 陈静仁. 大数据时代用户画像助力企业实现精准化营

销[J]. 中国集体经济, 2016(4): 61－62.
[10] 裴国才. 基于用户画像的电信精准营销模型研究[J]. 信息通讯, 2017(12): 240－241,24.
[11] 王晓霞, 刘静沙, 许丹丹. 运营商大数据用户画像实践[J]. 电信科学, 2018(5): 127－133.
[12] 盛怡瑾. 用户画像技术在学术期刊审稿人遴选中的应用[J]. 出版发行研究, 2018(8): 54－58.
[13] 熊伟, 杭波, 李兵, 等. 一种集成用户画像与内容的服务重定向方法[J]. 小型微型计算机系统, 2017(12): 2762－2765.
[14] 张诗军, 陈丰, 王志英, 等. 基于电力大数据的客户立体画像构建及应用研究[J]. 电气应用, 2018(8): 18－25.
[15] 高扬, 池雪花, 章成志, 等. 杰出人才精准画像构建研究——以智能制造领域为例[J]. 图书馆论坛, 2018(9): 1－8.
[16] 林艳霞, 谢湘生. 基于社会认同理论的微博群体用户画像[J]. 情报理论与实践,2017(3): 142－148.
[17] 陈添源. 高校移动图书馆用户画像构建实证[J]. 图书情报工作, 2018,62(7): 38－46.
[18] 陆冬磊. 基于电子商务的用户画像分析[J]. 电脑知识与技术, 2018(22): 1.
[19] 胡媛, 毛宁. 基于用户画像的数字图书馆知识社区用户模型构建[J]. 图书馆理论与实践,2017(4):82－85,97.
[20] 李冰, 王悦, 刘永祥. 大数据环境下基于 k-means 的用户画像与智能推荐的应用[J]. 现代计算机(专业版), 2016(24): 11－15.
[21] 张小可, 沈文明, 杜翠凤. 贝叶斯网络在用户画像构建中的研究[J]. 移动通讯, 2016(22): 22－26.
[22] 张钧. 基于用户画像的图书馆知识发现服务研究[J]. 图书与情报, 2017(6): 60－63.
[23] 王翔. 电影网站评分与电影票房关系研究——以豆瓣电影为例[D]. 南昌: 南昌大学, 2016.
[24] 薛欢雪. 高校图书馆学科服务用户画像创建过程[J]. 图书馆学研究, 2018(13): 67－71,82.
[25] 孙鸿飞, 武慧娟, 周兰萍. 基于标签的个性化信息推荐理论模型研究[J]. 情报科学, 2013, 31(4): 24－27.
[26] 杨晶, 成为青, 郭常忠. 基于标准标签的用户兴趣模型研究[J]. 计算机科学与技术, 2013(10): 208－211.
[27] 刘健. 数字图书馆资源聚合与服务推荐研究[D]. 长春: 吉林大学, 2017.

作者贡献说明：
徐海玲：论文撰写及修改；
张海涛：提出研究命题、研究思路；
魏明珠：英文文献的收集及摘要的翻译；
尹慧子：数据的采集及处理。

Research on the Construction of Social Media User Portrait and Resource Aggregation Model

Xu Hailing[1] Zhang Haitao[1,2] Wei Mingzhu[1] Yin Huizi[1]
[1] School of Management, Jilin University, Changchun 130022
[2] Information Resources Research Center, Jilin University, Changchun 130022

Abstract: [**Purpose/significance**] This paper creates a social media resource aggregation model based on user portraits, provides a reference for resource aggregation, and enriches and expands its theoretical research system. [**Method/process**] Based on the in-depth analysis of the connotation and algorithm of user portraits, and mapping relationship between them, the models of user portraits and resource portraits are constructed based on social media. The social media-based user portraits and resource portraits are studied by using the social labeling system method. Based on the domain ontology method, through the deep mining of data and the principle of resource aggregation, a social media internal resource aggregation model and an external resource aggregation model based on resource portraits are constructed. [**Result/conclusion**] In the era of big data, relevant theories and methods based on user portraits and resource portraits can provide new ideas for resource aggregation of social media.

Keywords: user portrait social media resource portrait resource aggregation